

11051
1600
RECEIVED

FEB 21 2002

OIPE

TECH CENTER 1600/2900

RAW SEQUENCE LISTING

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:10

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

SEQUENCE LISTING

#6/K.T.
7/2
Raw
Seq.

ENTERED

1 (1) GENERAL INFORMATION:

2 (i) APPLICANT: Tartaglia, Louis A.

3 Weng, Xun

4 (ii) TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING

5 GLUTEX AND USES THEREOF

6 (iii) NUMBER OF SEQUENCES: 10

7 (iv) CORRESPONDENCE ADDRESS:

8 (A) ADDRESSEE: Fish & Richardson P.C.

9 (B) STREET: 225 Franklin Street

10 (C) CITY: Boston

11 (D) STATE: MA

12 (E) COUNTRY: USA

13 (F) ZIP: 02110-2804

14 (v) COMPUTER READABLE FORM:

15 (A) MEDIUM TYPE: Diskette

16 (B) COMPUTER: IBM Compatible

17 (C) OPERATING SYSTEM: Windows95

18 (D) SOFTWARE: FastSEQ for Windows Version 2.0

19 (vi) CURRENT APPLICATION DATA:

C--> 20 (A) APPLICATION NUMBER: US/09/981,947A

C--> 21 (B) FILING DATE: 18-Oct-2001

23 (vii) PRIOR APPLICATION DATA:

24 (A) APPLICATION NUMBER: 09/031,392

25 (B) FILING DATE:

27 (viii) ATTORNEY/AGENT INFORMATION:

28 (A) NAME: Meiklejohn, Ph.D., Anita L.

29 (B) REGISTRATION NUMBER: 35,283

30 (C) REFERENCE/DOCKET NUMBER: 07334/072002

31 (ix) TELECOMMUNICATION INFORMATION:

32 (A) TELEPHONE: 617/542-5070

33 (B) TELEFAX: 617/542-8906

34 (C) TELEX: 200154

35 (2) INFORMATION FOR SEQ ID NO: 1:

36 (i) SEQUENCE CHARACTERISTICS:

37 (A) LENGTH: 2343 base pairs

38 (B) TYPE: nucleic acid

39 (C) STRANDEDNESS: single

40 (D) TOPOLOGY: linear

41 (ii) MOLECULE TYPE: cDNA

42 (ix) FEATURE:

43 (A) NAME/KEY: Coding Sequence

44 (B) LOCATION: 73...1761

RAW SEQUENCE LISTING

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:10

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

```

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
46 TCGACCCACG CGTCCGGCCT TGGCAGAGTC TGGGGTCCCT GGACTGAGCC ATCAGCTGGG 60
47 TCACTGAGAC CC ATG GCA AGG AAA CAA AAT AGG AAT TCC AAG GAA CTG GGC 111
48 Met Ala Arg Lys Gln Asn Arg Asn Ser Lys Glu Leu Gly
49 1 5 10
50 CTA GTT CCC CTC ACA GAT GAC ACC AGC CAC GCC GGG CCT CCA GGG CCA 159
51 Leu Val Pro Leu Thr Asp Asp Thr Ser His Ala Gly Pro Pro Gly Pro
52 15 20 25
53 GGG AGG GCA CTG CTG GAG TGT GAC CAC CTG AGG AGT GGG GTG CCA GGT 207
54 Gly Arg Ala Leu Leu Glu Cys Asp His Leu Arg Ser Gly Val Pro Gly
55 30 35 40 45
56 GGA AGG AGA AGA AAG GAC TGG TCC TGC TCG CTC CTC GTG GCC TCC CTC 255
57 Gly Arg Arg Arg Lys Asp Trp Ser Cys Ser Leu Leu Val Ala Ser Leu
58 50 55 60
59 GCG GGC GCC TTC GGC TCC TCC TTC CTC TAC GGC TAC AAC CTG TCG GTG 303
60 Ala Gly Ala Phe Gly Ser Ser Phe Leu Tyr Gly Tyr Asn Leu Ser Val
61 65 70 75
62 GTG AAT GCC CCC ACC CCG TAC ATC AAG GCC TTT TAC AAT GAG TCA TGG 351
63 Val Asn Ala Pro Thr Pro Tyr Ile Lys Ala Phe Tyr Asn Glu Ser Trp
64 80 85 90
65 GAA AGA AGG CAT GGA CGT CCA ATA GAC CCA GAC ACT CTG ACT CTG CTC 399
66 Glu Arg Arg His Gly Arg Pro Ile Asp Pro Asp Thr Leu Thr Leu Leu
67 95 100 105
68 TGG TCT GTG ACT GTG TCC ATA TTC GCC ATC GGT GGA CTT GTG GGG ACG 447
69 Trp Ser Val Thr Val Ser Ile Phe Ala Ile Gly Gly Leu Val Gly Thr
70 110 115 120 125
71 TTA ATT GTG AAG ATG ATT GGA AAG GTT CTT GGG AGG AAG CAC ACT TTG 495
72 Leu Ile Val Lys Met Ile Gly Lys Val Leu Gly Arg Lys His Thr Leu
73 130 135 140
74 CTG GCC AAT AAT GGG TTT GCA ATT TCT GCT GCA TTG CTG ATG GCC TGC 543
75 Leu Ala Asn Asn Gly Phe Ala Ile Ser Ala Ala Leu Leu Met Ala Cys
76 145 150 155
77 TCG CTC CAG GCA GGA GCC TTT GAA ATG CTC ATT GTG GGA CGC TTC ATC 591
78 Ser Leu Gln Ala Gly Ala Phe Glu Met Leu Ile Val Gly Arg Phe Ile
79 160 165 170
80 ATG GGC ATA GAT GGA GGC GTC GCC CTC AGT GTG CTC CCC ATG TAC CTC 639
81 Met Gly Ile Asp Gly Gly Val Ala Leu Ser Val Leu Pro Met Tyr Leu
82 175 180 185
83 AGT GAG ATC TCA CCC AAG GAG ATC CGT GGC TCT CTG GGG CAG GTG ACT 687
84 Ser Glu Ile Ser Pro Lys Glu Ile Arg Gly Ser Leu Gly Gln Val Thr
85 190 195 200 205
86 GCC ATC TTT ATC TGC ATT GGC GTG TTC ACT GGG CAG CTT CTG GGC CTG 735
87 Ala Ile Phe Ile Cys Ile Gly Val Phe Thr Gly Gln Leu Leu Gly Leu
88 210 215 220
89 CCC GAG CTG CTG GGA AAG GAG AGT ACC TGG CCA TAC CTG TTT GGA GTG 783
90 Pro Glu Leu Leu Gly Lys Glu Ser Thr Trp Pro Tyr Leu Phe Gly Val
91 225 230 235
92 ATT GTG GTC CCT GCC GTT GTC CAG CTG CTG AGC CTT CCC TTT CTC CCG 831
93 Ile Val Val Pro Ala Val Val Gln Leu Leu Ser Leu Pro Phe Leu Pro

```

RAW SEQUENCE LISTING

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:10

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

94	240	245	250	
95	GAC AGC CCA CGC TAC CTG CTC TTG GAG AAG CAC AAC GAG GCA AGA GCT	879		
96	Asp Ser Pro Arg Tyr Leu Leu Leu Glu Lys His Asn Glu Ala Arg Ala			
97	255 260 265			
98	GTG AAA GCC TTC CAA ACG TTC TTG GGT AAA GCA GAC GTT TCC CAA GAG	927		
99	Val Lys Ala Phe Gln Thr Phe Leu Gly Lys Ala Asp Val Ser Gln Glu			
100	270 275 280 285			
101	GTA GAG GAG GTC CTG GCT GAG AGC CAC GTG CAG AGG AGC ATC CGC CTG	975		
102	Val Glu Glu Val Leu Ala Glu Ser His Val Gln Arg Ser Ile Arg Leu			
103	290 295 300			
104	GTG TCC GTG CTG GAG CTG CTG AGA GCT CCC TAC GTC CGC TGG CAG GTG	1023		
105	Val Ser Val Leu Glu Leu Leu Arg Ala Pro Tyr Val Arg Trp Gln Val			
106	305 310 315			
107	GTC ACC GTG ATT GTC ACC ATG GCC TGC TAC CAG CTC TGT GGC CTC AAT	1071		
108	Val Thr Val Ile Val Thr Met Ala Cys Tyr Gln Leu Cys Gly Leu Asn			
109	320 325 330			
110	GCA ATT TGG TTC TAT ACC AAC AGC ATC TTT GGA AAA GCT GGG ATC CCT	1119		
111	Ala Ile Trp Phe Tyr Thr Asn Ser Ile Phe Gly Lys Ala Gly Ile Pro			
112	335 340 345			
113	CCG GCA AAG ATC CCA TAC GTC ACC TTG AGT ACA GGG GGC ATC GAG ACT	1167		
114	Pro Ala Lys Ile Pro Tyr Val Thr Leu Ser Thr Gly Gly Ile Glu Thr			
115	350 355 360 365			
116	TTG GCT GCC GTC TTC TCT GGT TTG GTC ATT GAG CAC CTG GGA CGG AGA	1215		
117	Leu Ala Ala Val Phe Ser Gly Leu Val Ile Glu His Leu Gly Arg Arg			
118	370 375 380			
119	CCC CTC CTC ATT GGT GGC TTT GGG CTC ATG GGC CTC TTC TTT GGG ACC	1263		
120	Pro Leu Leu Ile Gly Gly Phe Gly Leu Met Gly Leu Phe Gly Thr			
121	385 390 395			
122	CTC ACC ATC ACG CTG ACC CTG CAG GAC CAC GCC CCC TGG GTC CCC TAC	1311		
123	Leu Thr Ile Thr Leu Thr Leu Gln Asp His Ala Pro Trp Val Pro Tyr			
124	400 405 410			
125	CTG AGT ATC GTG GGC ATT CTG GCC ATC ATC GCC TCT TTC TGC AGT GGG	1359		
126	Leu Ser Ile Val Gly Ile Leu Ala Ile Ile Ala Ser Phe Cys Ser Gly			
127	415 420 425			
128	CCA GGT GGC ATC CCG TTC ATC TTG ACT GGT GAG TTC TTC CAG CAA TCT	1407		
129	Pro Gly Gly Ile Pro Phe Ile Leu Thr Gly Glu Phe Phe Gln Gln Ser			
130	430 435 440 445			
131	CAG CGG CCG GCT GCC TTC ATC ATT GCA GGC ACC GTC AAC TGG CTC TCC	1455		
132	Gln Arg Pro Ala Ala Phe Ile Ile Ala Gly Thr Val Asn Trp Leu Ser			
133	450 455 460			
134	AAC TTT GCT GTT GGG CTC CTC TTC CCA TTC ATT CAG AAA AGT CTG GAC	1503		
135	Asn Phe Ala Val Gly Leu Leu Phe Pro Phe Ile Gln Lys Ser Leu Asp			
136	465 470 475			
137	ACC TAC TGT TTC CTA GTC TTT GCT ACA ATT TGT ATC ACA GGT GCT ATC	1551		
138	Thr Tyr Cys Phe Leu Val Phe Ala Thr Ile Cys Ile Thr Gly Ala Ile			
139	480 485 490			
140	TAC CTG TAT TTT GTG CTG CCT GAG ACC AAA AAC AGA ACC TAT GCA GAA	1599		
141	Tyr Leu Tyr Phe Val Leu Pro Glu Thr Lys Asn Arg Thr Tyr Ala Glu			
142	495 500 505			

RAW SEQUENCE LISTING

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:10

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

```

143   ATC AGC CAG GCA TTT TCC AAA AGG AAC AAA GCA TAC CCA CCA GAA GAG      1647
144   Ile Ser Gln Ala Phe Ser Lys Arg Asn Lys Ala Tyr Pro Pro Glu Glu
145   510                               515                               520                               525
146   AAA ATC GAC TCA GCT GTC ACT GAT GCT CCT GCT TCT TCT CCT TTC ACT      1695
147   Lys Ile Asp Ser Ala Val Thr Asp Ala Pro Ala Ser Ser Pro Phe Thr
148                                 530                               535                               540
149   ACT CCG AAT ACA GCC TGG ATT CAA GCT GCC GCC ACC ACC ACC GCC ACC      1743
150   Thr Pro Asn Thr Ala Trp Ile Gln Ala Ala Thr Thr Thr Ala Thr
151                                 545                               550                               555
152   AAA AAA GAA CAC CCA TTG TAAACGGTCA TGTGGTATTT CCTCAACCTG GAATGACC      1799
153   Lys Lys Glu His Pro Leu
154   560
155   TTCCCTATC TTCTTCTCCT GGAGAACACC AAGTCATGAT GTCAGACAAG AGCTTGGATT      1859
156   TTGGAGACAT GGGTTTGAAT TCCAGTCATT CATTTCTTTTA TTCAGCAAAT ATTTAACAAG      1919
157   TACTGACATG TCCCATATGT TGTTTTACCC ACTGGTTATA CAATGGGAGG GAGAGAGAGA      1979
158   GAGAGAGAGA GAGAGAGATG CTATTCTAAA AGCTTGAAGT CTAGGCTGTG CACGGTGGCT      2039
159   CACGCCTGTA ATCCCAGCAC TTTGGGAGGC CGAGGTGGGT GGATCGTGAG GTCAGGAGAT      2099
160   TGAGACCATC CTGGCTAACA TGGTGAAACT CCCTCTCTAC TAAAAATACA AAAAATTAGC      2159
161   TGAGCATGGT GGC GGCGGCC TGTAGTCCCA GCTACTTGGG AGGCTGAGGC AGGAGAATGG      2219
162   CGTGAACCCA GGAGGCGGAG CTTGCAGTGA GCCGAGATCA CACCACCACA CTCCAGCCTG      2279
163   GGTGACAGAG CCAGACTCCG TCTCAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAGGGCGG      2339
164   CCGC                                                                2343

```

166 (2) INFORMATION FOR SEQ ID NO: 2:

167 (i) SEQUENCE CHARACTERISTICS:

168 (A) LENGTH: 563 amino acids

169 (B) TYPE: amino acid

170 (D) TOPOLOGY: linear

171 (ii) MOLECULE TYPE: protein

172 (v) FRAGMENT TYPE: internal

173 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

174   Met Ala Arg Lys Gln Asn Arg Asn Ser Lys Glu Leu Gly Leu Val Pro
175   1           5           10           15
176   Leu Thr Asp Asp Thr Ser His Ala Gly Pro Pro Gly Pro Gly Arg Ala
177   20           25           30
178   Leu Leu Glu Cys Asp His Leu Arg Ser Gly Val Pro Gly Gly Arg Arg
179   35           40           45
180   Arg Lys Asp Trp Ser Cys Ser Leu Leu Val Ala Ser Leu Ala Gly Ala
181   50           55           60
182   Phe Gly Ser Ser Phe Leu Tyr Gly Tyr Asn Leu Ser Val Val Asn Ala
183   65           70           75           80
184   Pro Thr Pro Tyr Ile Lys Ala Phe Tyr Asn Glu Ser Trp Glu Arg Arg
185   85           90           95
186   His Gly Arg Pro Ile Asp Pro Asp Thr Leu Thr Leu Leu Trp Ser Val
187   100          105          110
188   Thr Val Ser Ile Phe Ala Ile Gly Leu Val Gly Thr Leu Ile Val
189   115          120          125
190   Lys Met Ile Gly Lys Val Leu Gly Arg Lys His Thr Leu Leu Ala Asn
191   130          135          140
192   Asn Gly Phe Ala Ile Ser Ala Ala Leu Leu Met Ala Cys Ser Leu Gln

```

RAW SEQUENCE LISTING

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:10

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

193	145	150	155	160
194	Ala Gly Ala Phe Glu Met Leu Ile Val Gly Arg Phe Ile Met Gly Ile			
195		165	170	175
196	Asp Gly Gly Val Ala Leu Ser Val Leu Pro Met Tyr Leu Ser Glu Ile			
197		180	185	190
198	Ser Pro Lys Glu Ile Arg Gly Ser Leu Gly Gln Val Thr Ala Ile Phe			
199		195	200	205
200	Ile Cys Ile Gly Val Phe Thr Gly Gln Leu Leu Gly Leu Pro Glu Leu			
201		210	215	220
202	Leu Gly Lys Glu Ser Thr Trp Pro Tyr Leu Phe Gly Val Ile Val Val			
203		225	230	235
204	Pro Ala Val Val Gln Leu Leu Ser Leu Pro Phe Leu Pro Asp Ser Pro			
205		245	250	255
206	Arg Tyr Leu Leu Leu Glu Lys His Asn Glu Ala Arg Ala Val Lys Ala			
207		260	265	270
208	Phe Gln Thr Phe Leu Gly Lys Ala Asp Val Ser Gln Glu Val Glu Glu			
209		275	280	285
210	Val Leu Ala Glu Ser His Val Gln Arg Ser Ile Arg Leu Val Ser Val			
211		290	295	300
212	Leu Glu Leu Leu Arg Ala Pro Tyr Val Arg Trp Gln Val Val Thr Val			
213		305	310	315
214	Ile Val Thr Met Ala Cys Tyr Gln Leu Cys Gly Leu Asn Ala Ile Trp			
215		325	330	335
216	Phe Tyr Thr Asn Ser Ile Phe Gly Lys Ala Gly Ile Pro Pro Ala Lys			
217		340	345	350
218	Ile Pro Tyr Val Thr Leu Ser Thr Gly Gly Ile Glu Thr Leu Ala Ala			
219		355	360	365
220	Val Phe Ser Gly Leu Val Ile Glu His Leu Gly Arg Arg Pro Leu Leu			
221		370	375	380
222	Ile Gly Gly Phe Gly Leu Met Gly Leu Phe Phe Gly Thr Leu Thr Ile			
223		385	390	395
224	Thr Leu Thr Leu Gln Asp His Ala Pro Trp Val Pro Tyr Leu Ser Ile			
225		405	410	415
226	Val Gly Ile Leu Ala Ile Ile Ala Ser Phe Cys Ser Gly Pro Gly Gly			
227		420	425	430
228	Ile Pro Phe Ile Leu Thr Gly Glu Phe Phe Gln Gln Ser Gln Arg Pro			
229		435	440	445
230	Ala Ala Phe Ile Ile Ala Gly Thr Val Asn Trp Leu Ser Asn Phe Ala			
231		450	455	460
232	Val Gly Leu Leu Phe Pro Phe Ile Gln Lys Ser Leu Asp Thr Tyr Cys			
233		465	470	475
234	Phe Leu Val Phe Ala Thr Ile Cys Ile Thr Gly Ala Ile Tyr Leu Tyr			
235		485	490	495
236	Phe Val Leu Pro Glu Thr Lys Asn Arg Thr Tyr Ala Glu Ile Ser Gln			
237		500	505	510
238	Ala Phe Ser Lys Arg Asn Lys Ala Tyr Pro Pro Glu Glu Lys Ile Asp			
239		515	520	525
240	Ser Ala Val Thr Asp Ala Pro Ala Ser Ser Pro Phe Thr Thr Pro Asn			
241		530	535	540

VERIFICATION SUMMARY

DATE: 02/08/2002

PATENT APPLICATION: US/09/981,947A

TIME: 10:08:11

Input Set : N:\Crf3\RULE60\09981947.raw

Output Set: N:\CRF3\02082002\I981947A.raw

L:20 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:21 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:598 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=8
L:608 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=9
L:619 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:625 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:629 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:633 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:649 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:651 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:677 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10
L:679 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10